Accepted author version, in press at *Cortex*, doi: 10.1016/j.cortex.2020.09.014

Facing Dyads and Single Faces in the Social Visual World

Commentary on Papeo (in press), Twos in human visual perception.

H. Wiese

Department of Psychology, Durham University, United Kingdom

In the viewpoint article by Papeo published in this issue of Cortex, the author reviews experimental work on the relevance of so-called facing dyads, i.e., two bodies facing each other, relative to non-facing body pairs for visual processing. Quoting Ken Nakayama (2011) on how vision has been "shaped by the requirements of social life" (p.3), the author suggests that the visual system is specifically tuned to process such facing dyads, as they indicate a social interaction and therefore particular saliency. Papeo (in press) presents important and interesting work from a relatively new research area, which will undoubtedly stimulate and impact future studies. To highlight potential paths for further empirical research and theoretical integration, this commentary aims at viewing the discussed findings both in the broader context of person perception and in the more specific context of face perception. I will focus on three aspects, namely (i) on person perception as the integration of information from different input domains, (ii) on inversion effects and configural processing in body and face perception, as well as (iii) on the question of domain specificity.

Person perception as the integration of information from different input domains

Broadly in line with Nakayama (2011), it appears reasonable to assume that the ultimate aim of social vision is to understand and successfully navigate the world around us,

which is filled with other people pursuing their own personal or group goals. For this aim, it is essential not only (i) to detect, categorise and recognise the stimuli that relate to these social agents, but also (ii) to access knowledge about the perceived people and groups, and (iii) to predict their behaviour and plan our own actions. While mnemonic and planning processes are involved in the latter two later stages, social vision might be seen as reflecting the first step in this sequence.

The research discussed by Papeo (in press) describes the initial part of this first visual processing step, namely "facing-dyad detection", or the process of perceiving that a visual stimulus represents a facing dyad as opposed to a different visual object. Given the sequential architecture of person perception, this first step is crucial in the sense that all further processing depends on it. However, as necessary as this first step is, it is arguably not sufficient for social vision. Once a facing dyad is detected, it will be crucial to find out who the two involved people are and what they are doing (or what they are likely to do next). Surely, it will often be of considerable relevance whether the detected dyad is fighting or whether the two people are collaboratively working together. Similarly, the meaning of the perceived stimulus will change depending on whether we see two friends or two competitors collaborating on a task.

Accordingly, for social vision to be successful, the observer needs to find out as much as possible about the perceived person(s). In face research, a number of cognitive and neuroscientific models have been proposed to describe this process of information accumulation (Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000; Schweinberger & Burton, 2003), and it appears that familiarity is particularly important for guiding further processing (Young & Burton, 2017). If the face is familiar, the specific individual can usually be recognised and identity-specific semantic, episodic and affective information can be accessed (see e.g. Wiese et al., 2019). If the face is unfamiliar, individual identity is

unknown, but the observer will usually still be able to categorise the face on a number of important dimensions, such as gender, relative age, and ethnic background. All this information is crucial for social interaction, as we talk differently to close friends relative to strangers, and we interact differently with an adolescent relative to an older adult person.

While all this information is typically readily accessible from the face, it seems more difficult, or in some cases even impossible, to infer it from the perception of the body alone. Crucially, a detection mechanism for facing dyads as discussed by Papeo (in press) may act as a trigger for processing more detailed information from the face(s), and it seems plausible that body- and face-sensitive areas in the brain interact to achieve this. Interestingly, such integration of different stimulus domains seems to be shown by an fMRI experiment discussed in the paper (Abassi & Papeo, 2020), in which the so-called Fusiform Face Area was found to be more responsive to facing dyads relative to non-facing dyads. A close neural coupling of input systems for person perception seems plausible, as the perception of a body is predictive of the perception of a face in the real world. An interesting question then for future research is how exactly these different systems of person perception are integrated.

Inversion effects and configural processing

Broadly in line with the above described integration, Papeo (in press) suggests analogies of processing facing dyads with face perception, and a number of empirical findings are discussed to support this suggestion. First, facing dyads show what the author calls the two-body inversion effect (Papeo, Stein, & Soto-Faraco, 2017). In other words, rotating stimuli by 180 degrees disrupts the detection of facing bodies more than non-facing dyads when participants have to categorize the stimuli as either bodies or objects. This inversion effect is interpreted as an index for the “visual tuning to a unitary configuration of two bodies facing each other as if interacting” (p.5).

Face inversion effects are typically interpreted as reflecting interference with configural and holistic processing (e.g., Maurer, Grand, & Mondloch, 2002). The manipulation is known to disrupt the processing of identity, i.e., it hampers the recognition of an individual face. However, it also impairs face detection, and this appears to be where the analogy to the two-body inversion effect is located. We perceive a stimulus as a face, or as face-like, when we see two eyes above a nose above a mouth, and inversion disrupts this basic configuration. Similarly, the impaired detection of facing dyads in inverted stimuli may be related to difficulties of processing critical spatial relations.

As will necessarily be the case in a relatively young area of research, a number of questions remain unanswered. First, it would be interesting to examine why the inversion effect does not generalise to human-object pairs. The author seems to suggest that the specific social nature of two bodies potentially interacting with each other triggers the effect because of its high "biological relevance" to the viewer (p.11). Arguably, however, a single person manipulating an object could be highly relevant as well. Second, the control condition for two bodies facing each other appears to be two bodies turned 180 degrees away from each other. Clearly, one of the two bodies facing, and therefore potentially interacting with, the observer would also indicate high relevance. Future studies may ask what exactly is driving these inversion effects, and exactly which visual elements are necessary to trigger it.

As noted above, the perceptual process discussed here reflects the *detection* of a specific visual stimulus, and the processing of a basic face-like configuration seems indeed necessary for this detection process in face perception. At the same time, there is now abundant evidence that the *recognition* of individual identity in faces is not strongly related to "subtle variations in shape and spatial relations between parts" (p.12; see Burton, Schweinberger, Jenkins, & Kaufmann, 2015), or second-order configural information. Instead, familiar face recognition seems to rely more strongly on surface-reflectance and

texture cues. It again appears that the visual system is tuned to detect specific potentially socially relevant stimuli, such as facing dyads or faces, and that this early detection mechanism is then followed by a sequence of different processes that are equally necessary to establish the social meaning of these stimuli.

Facing dyads as "face-like special" visual stimuli

In face research, the "specificity question" is nearly as old as the field itself. It seems that whenever a researcher claims that faces are (or are not) "special" relative to other visual objects, it will be discussed anew in one or the other form. In the present viewpoint article, Papeo (in press) concludes that facing dyads are "face-like special". This conclusion is based on similar experimental effects (such as the inversion effects discussed above) and therefore similar neural and visual tuning.

Independent of whether the empirical evidence for this claim, or the general idea of domain specificity in higher-level visual processing, is seen as convincing, it might be interesting to ask why exactly facing dyads would be special, particularly relative to other visual stimuli involving a person. Papeo (in press) suggests that facing dyads are "biologically relevant" (p.11) but does not go into detail as to how. As already noted, it appears easy to construct a scenario in which a single person manipulating an object might be even more relevant than two people facing each other (e.g., somebody picking up a weapon s/he could use against us). Moreover, it seems that we often see facing dyads that are not particularly relevant (e.g. on a crowded bus or train platform). It therefore appears that additional empirical and theoretical work will be helpful to understand fully the meaning of the findings discussed in this article.

In conclusion, the research presented in this viewpoint article brings up a number of interesting yet unanswered questions, and given the recency of the discussed work this will

almost inevitably be the case. I am very much looking forward to seeing how this field of person perception will develop in the years to come.

References

Abassi, E., & Papeo, L. (2020). The Representation of Two-Body Shapes in the Human Visual Cortex. *J Neurosci, 40*(4), 852-863. doi:10.1523/JNEUROSCI.1378-19.2019

Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3)*, 305-327. doi:10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments Against a Configural Processing Account of Familiar Face Recognition. *Perspect Psychol Sci, 10*(4), 482-496. doi:10.1177/1745691615583129

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci, 4*(6), 223-233. doi:10.1016/s1364-6613(00)01482-0

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cogn Sci, 6*(6), 255-260. doi:10.1016/s1364-6613(02)01903-4

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The Two-Body Inversion Effect. *Psychol Sci, 28*(3), 369-379. doi:10.1177/0956797616685769

Schweinberger, S. R., & Burton, A. M. (2003). Covert recognition and the neural system for face processing. *Cortex, 39*(1), 9-30. doi:10.1016/s0010-9452(08)70071-6

Wiese, H., Tuttenberg, S. C., Ingram, B. T., Chan, C. Y. X., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019). A Robust Neural Index of High Face Familiarity. *Psychological Science, 30*(2), 261-272. doi:10.1177/0956797618813572

Young, A. W., & Burton, A. M. (2017). Recognizing Faces. *Current Directions in Psychological Science, 26*(3), 212-217. doi:10.1177/0963721416688114